R E P O R T R E S U M E S

ED 013 460 CG 000 745

THE EFFECT OF FEEDBACK TRAINING ON ACCURACY OF JUDGMENTS.

BY- WATLEY, DONIVAN J.
NATIONAL MERIT SCHOLARSHIP CORP., EVANSTON, ILL.

REPORT NUMBER NMSC-RR-VOL-3-NO-3-1967 PUB DATE 67
EDRS PRICE MF-$0.25 HC-$0.68 17P.

DESCRIPTORS- *COLLEGE STUDENTS, MALES, *GRADE PREDICTION, RESEARCH, *TRAINING, *COUNSELORS, EXPECTANCY TABLES, COLLEGE FRESHMEN, TEST RESULTS, F TEST, STRONG VOCATIONAL INTEREST BLANK, SEQUENTIAL TEST OF EDUCATIONAL PROGRESS, T TEST, MINNESOTA SCHOLASTIC APTITUDE TEST, COOPERATIVE ENGLISH TEST

OPINIONS DIFFER ABOUT THE BEST METHOD FOR TRAINING JUDGES TO MAKE CLINICAL FORECASTS. SOME EVIDENCE SUGGESTS, HOWEVER, THAT JUDGMENTS ARE MORE LIKELY TO IMPROVE UNDER PREDICTION CONDITIONS THAT ARE PRECISELY DEFINED. THIS STUDY ASSESSED THE EFFECT OF PROVIDING IMMEDIATE FEEDBACK TRAINING TO JUDGES KNOWN FROM A PREVIOUS STUDY TO PREDICT EDUCATIONAL CRITERIA AT RELATIVELY HIGH, MODERATE, OR LOW LEVELS OF ACCURACY. THE CRITERIA PREDICTED WERE FRESHMAN AND OVERALL COLLEGE GRADES. IN COMPARISON WITH JUDGES WHO RECEIVED NO TRAINING, THE FORECASTS OF "LOW" ACCURACY JUDGES SHOWED SUBSTANTIAL IMPROVEMENTS FOR BOTH PREDICTED CRITERIA. HOWEVER, THE TRAINING HAD NO NOTICEABLE AFFECT ON THE JUDGEMENTS OF THE "HIGH" OR "MODERATE" ACCURACY JUDGES. THIS DOCUMENT IS A NATIONAL MERIT SCHOLARSHIP CORPORATION REPORT, VOLUME 3, NUMBER 3, 1967. (AUTHOR)

NMSC

ED013460

1967: volume 3, number 3

The Effect of Feedback Training on Accuracy of Judgments

Donivan J. Watley

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

NATIONAL MERIT SCHOLARSHIP CORPORATION

CG 000 745

Abstract

Opinions differ about the best method for training judges to make clinical forecasts. Some evidence suggests, however, that judgments are more likely to improve under prediction conditions that are precisely defined. This study assessed the effect of providing immediate feedback training to judges known from a previous study to predict educational criteria at relatively high, moderate, or low levels of accuracy. The criteria predicted were freshman and overall college grades. In comparison with judges who received no training, the forecasts of "low" accuracy judges showed substantial improvements for both predicted criteria; however, the training had no noticeable affect on the judgments of the "high" or "moderate" accuracy judges.

The Effect of Feedback Training on Accuracy of Judgments

Donivan J. Watley

National Merit Scholarship Corporation, Evanston, Illinois

Although psychologists frequently engage in attempts to forecast behavior, the accuracy of clinical judgment leaves much to be desired (Meehl, 1954; Goldman, 1961; Gough, 1962; Oskamp, 1962; and Sawyer, 1966). Two general views have been proposed about the best method of designing training for the purpose of making clinical forecasts. McArthur (1954) argued that students should begin with global data like autobiographies and dream material in order to emphasize the understanding of personality themes and case theories, with the usual basic identifying test data and sterotyped information avoided in the early stages because this seems to have the effect of freezing the judge's impression of the "person." The implicit, though as yet unsupported, assumption is that a more fully completed case theory enables the judge to make more accurate forecasts. On the other hand, Cronbach (1956), Meehl (1954), Tyler (1961), and Goldman (1961) maintain that students should concentrate heavily on learning stereotypes, base-rates, and averages. They argue that the available empirical evidence clearly justifies that minimum inference by the judge is the best approach to making predictions.

The best approach to training clinical judges is obviously complex, and exponents of the two approaches seem to have somewhat different goals in mind. The approach that promotes minimum inference in prediction is primarily concerned with outcome, institutional type forecasts (e.g., college grades; accept or reject for therapy) where the relationships between well-defined variables and concisely defined criteria have been determined. Considerable

evidence has accumulated showing that clinical judges cannot outpredict mechanical, straightforward predictions in these situations. Thus, training for this type of forecast is a matter of acquainting judges with base-rates, averages, etc., and avoiding inferences that tend to decrease accuracy. But McArthur's comments were mostly directed toward training skillful and insightful therapists who see beyond a few pieces of case data. To "understand" the client, the trainee is encouraged to make inferences in developing a case theory, and this process may or may not proceed to making outcome or institutional type predictions. There is a basic difference in <u>understanding</u> the client and making institutional-type forecasts <u>about</u> the individual. McArthur may be incorrect, however, in assuming that all "... good predictions ... come from the construct as a whole" (1954, p. 204) rather than from one or two pieces of case data.

Taft's (1955 and 1959) reviews provide some clues about the feasibility of training to improve the accuracy of clinical judgments. He noted that psychologists have tended to predict best when technical tasks were involved and that predictive ability seems more likely to improve when the training provided is specific. He also suggested that psychologists' judgments are negatively affected by an overconcern with attempts to perceive individual differences.

Taft's conclusions are supported by the work of several investigators. Both Cline (1955) and Oskamp (1962) demonstrated that "experts" predicted significantly better than students when specific training programs were used in addition to clearly defined criteria. Using judgments of psychiatric hospitalization versus nonpsychiatric (medical) hospitalization, Oskamp concluded that clinical training programs might profitably use objective predictive tasks accompanied by specific training and immediate feedback to speed the development of internal norms; and that this method may be preferable to the more common technique of prolonged and

intensive analysis of a few cases. On the other hand, Crow (1957) found that training in responding to individual differences and the abandoning of stereotypes decreased the accuracy of interpersonal perception, which apparently could be attributed to a more-than-optimal increase in sensitivity to individual differences; Watley (1967) found evidence that predictive accuracy decreased slightly for some judges after they were given considerable information to integrate for themselves that, synthesized and used appropriately, should have helped improve judgments of college grades; and Soskin (1954) reported that general training in the use of projective test information did not improve the accuracy of clinical forecasts.

Thus, although there are advantages to both the minimum inference and maximum inference approaches to training, the evidence so far favors the former method when prediction tasks are involved. The purpose of this study was to determine whether the judgments of educational counselors could be improved for the well-defined criterion of college grades. Previous research (Watley, 1966b) showed that counselors vary greatly in their ability to predict this criterion, with many unable to use case information effectively. Immediate feedback data were provided primarily for the purpose of developing internal norms and helping judges become more aware of specific variables to emphasize in making predictions. The effectiveness of this type of training was assessed by determining the gains in accuracy of judges known from an earlier study (Watley, 1966b) to predict at relatively high, moderate, or low levels of accuracy.

Method

Judges

Thirty-six counselors took part in this study, all of whom were in a previous study (Watley, 1966b) that assessed individual differences in predictive

ability. The initial study included 66 judges and selection for the present study was based on their ability to predict: (1) freshman grades, (2) overall college grades, and (3) whether students would persist and be successful in the educational programs they selected at the time of admission to college.

Using their prediction records, the judges were ranked from 1 to 66 on each of the three criteria. The two ranks for freshman and overall college grades were then combined, leaving one set of ranks for accuracy in forecasting grades and the other for judging persistence and graduation from initial educational programs. Twelve judges were identified who ranked in the top one-third (including ranks 1-22), 12 in the middle one-third (ranks 23-44), and 12 in the bottom one-third (ranks 45-66) on <u>each</u> of the two sets of rankings. Judges at the three levels were labeled, respectively, the high, moderate, and low accuracy groups.

No differences were found among the high, moderate, and low accuracy groups on the amount of counseling experience accrued. The high group had on the average only slightly more counselor training than the other groups.

<u>Training</u>

This investigation was conducted approximately one year after the initial study (Watley, 1966b). Each judge was given information about the number of hits (correct "C or better" or "less than C" judgments) he recorded in the first study and the correlation coefficient between his predictions and the grades actually earned by students. In addition, information was provided about the case variables most highly related to the predicted criteria (freshman and overall college grades), as well as the difference in data typically used by judges who predict at relatively high, moderate, or low levels of accuracy. Data were also given about: the relationship between counselor confidence in

their judgments and actual predictive accuracy (Watley, 1966b); the effect of place of employment (high school or college) on counselor judgments; and psychometric and biometric differences among counselors who predicted educational criteria most or least accurately. However, no specific training was involved in this phase of the study; the judges were only provided with information obtained in the initial study.

An attempt was made, however, to determine whether the prediction experience acquired in the initial study (Watley 1966b) plus the information provided in the first phase of this study significantly affected the accuracy of counselor judgments. No evidence was obtained that this type of experience improved forecasts, and some evidence showed that the predictions of some judges were slightly less accurate (Watley, 1967).

The next phase of this study involved the specific feedback training. Six of the 12 judges in each of the high, moderate, and low accuracy groups were randomly selected to receive the training, with the other six in each group making predictions without receiving any further training. The effect of the training program was determined by comparing their judgments with those of the untrained groups when the total amount of prediction experience for all groups was approximately equal.

The training consisted of immediate feedback after freshman and overall college grade judgments were made for each case. After making his judgments for a particular case, the judge immediately received the following information: the student's actual freshman and overall college grades; the student's complete college grade transcript (at the University of Minnesota); statistically predicted freshman and overall grades using High School Rank (HSR), the Minnesota Scholastic Aptitude Test (MSAT), and the Cooperative English Test (CET) as predictor variables; and a complete list of the student's changes in educational major. The information provided was factual, and it was kept to a minimum in order to

enhance the judge's efforts toward effective integration. The judge was allowed to study the feedback data as long as he wished before proceeding to the next case. This procedure was followed for all 50 cases by the trained group. The untrained group made judgments for the same 50 cases but received no feedback data.

Prediction Sample and Case Data

The sample was composed of 50 males who entered the College of Science, Literature, and the Arts (SLA) at the University of Minnesota as first quarter freshman in the fall of 1959. They were randomly selected from among the entire entering class of freshman males. However, inclusion in this study depended on the availability of all of the desired psychometric and biographic case data, graduation from a Minnesota high school during the spring of 1959, and at least one quarter spent in SLA.

Information about scholastic aptitude and past academic achievement was given in a folder containing all of the data compiled for each student. Test scores were provided for the MSAT, the CET, and the Social Studies Test of the Sequential Tests of Educational Progress. Achievement data included each student's HSR and the last high school grades earned in the areas of mathematics, English, social studies, and natural sciences. Also included were results for the Strong Vocational Interest Blank and the Minnesota Multiphasic Personality Inventory, plus considerable biographic information given on the Minnesota College Admissions Form and the Personal Inventory for entering students.

Statistical data were also provided to each judge for use in making forecasts. This included: freshman grade expectancy tables for HSR and the MSAT; zero-order and multiple correlations between freshman grades and HST, MSAT, and the CET; and a regression equation that included prediction coefficients for the high school grades of mathematics, English, social studies, and natural sciences.

The type and amount of case information provided in these folders corresponded to the third condition under which judgments were made in the initial study (Watley, 1966b). These folders contained essentially all of the data that were available for this group of students before they entered college.

Results

Summary data for the trained and untrained judges are shown in Table 1. An analysis of variance was computed separately for each of the two predicted criteria.

Table 1

Summary Data for the Mean Number of Hits Obtained by Trained and Untrained Judges

Group		Level of Predictive Skill					
		High		Moderate		Low	
		First Year	O-A	First Year	O-A	First Year	O-A
Untrained	Mn	36.8	30.5	32.5	27.2	29.0	27.8
	SD	1.6	2.0	5.8	2.6	5.8	2.8
Trained	Mn	37.0	30.5	31.7	28.7	35.0	32.7
	SD	2.2	2.6	6.0	4.3	1.9	3.0

Note.--Predicted Criteria: first year grade average and overall grade average (O-A).

For freshman grades, the obtained F-ratio of 1.33 for assessing the total mean difference between the trained and untrained counselors was not significant at the .05 level. Thus, the specific feedback training did not appear to have the effect of generally improving the accuracy of judgments when the initial level of predictive ability was not taken into account. Yet, the high accuracy

group was already predicting close to the highest level of accuracy currently possible, with their predictions about as accurate as those made by the statistical method (Watley, 1966b). But in contrast, the low accuracy judges might be expected to gain more from the training experience since they started at an accuracy level below that of the equation or the best judges. Table 1 shows that the mean hits for the high accuracy trained and untrained judges were virtually the same; and, likewise, the moderate accuracy trained and untrained groups made similar hit records. However, the low accuracy trained judges averaged six more hits than the untrained judges, a difference that is significant at the .05 level. The judges who initially predicted at the lowest level of accuracy benefited from the training by improving their predictive accuracy to a level comparable to the judges who predicted at the highest level of accuracy.

The F of 8.87 for assessing the total mean differences among the high, moderate, and low accuracy groups was significant beyond the .01 level. Except for the low accuracy trained group, the experiences provided in this study failed to have any noticeable effect on the relative efficiency of the three accuracy groups. The interaction term was not significant at the .05 level.

For overall college grades, the F of 4.04 for assessing the total mean difference between the trained and untrained counselors was not significant at the .05 level. Thus, as with freshman grades, the training provided in this study failed to generally improve the accuracy of judgments when the initial level of predictive ability was not controlled. While no differences were observed between the high accuracy trained and untrained judges or between the moderate accuracy trained and untrained groups, the low accuracy trained judges made on the average about five more hits than the untrained low accuracy judges. The t for this difference was significant at the .05 level. Table 1 shows that the low accuracy trained group actually exceeded the mean number of hits obtained by the high accuracy trained judges.

The F of 4.91 for assessing the total mean differences between the three accuracy groups was significant at the .05 level. As was found with the freshman grade predictions, only the low accuracy trained group showed any relative improvement. The interaction term was not significant at the .05 level.

Discussion

This study attempted to provide information about training judges in making forecasts of specific educational criteria. The results are of most value, however, when they are interpreted in light of other findings of earlier studies of this series.

In the initial study (Watley, 1966 b), it was found that judges varied markedly in their ability to accurately predict college grades. Whereas some judges predicted about as accurately as the statistical equation normally used to forecast grades for the sample of students studied, others were unable to predict better than the pass-fail baserate. In an effort to improve their forecasting skills, judges were then given information about their predictive performance in the first study, about variables most highly correlated with the criterion, about variables typically used by judges who predict most or least accurately, and other general information that could have been used to improve predictive accuracy (Watley, 1967). It is important to note, however, that no specific training program was involved. Rather, the counselors were left with the job of integrating and synthesizing this information for themselves. Predictive accuracy did not improve under these conditions and some evidence was found that accuracy slightly decreased for some judges. But these results are similar to those found in other studies (e.g., Crow, 1957;

Soskin, 1954) where clinical forecasts failed to improve when the relationship between predictors and criteria was inadequately defined or poorly understood by the judges. Also, these findings are not supportive of McArthur's assumption that judges can effectively integrate and evaluate data to produce meaningful associations useful for predicting outcome criteria. However, perhaps a distinction needs to be made. Though unable to accurately forecast outcome-type criteria, the clinical judge may still be able to accurately describe the client and he may be highly accurate with the particular forecasts he chooses to make (e.g., "Your working under direct, dominant authority will probably produce this form of behavior").

The present study used predictors and criteria that were precisely defined, and judges were given specific feedback information immediately following each judgment. With this procedure, the judges had the opportunity to evaluate each forecast and make immediate adjustments. Still, judges differed initially in their predictive ability and, thus, some had more room for improvement than others. In this study the best judges did not improve their forecasts, but they already predicted close to the highest level of accuracy currently possible. But the judges who initially predicted least accurately improved their judgments to a level equal to that of the most accurate judges. They made similar improvement on both predicted criteria--freshman and overall college grades.

These results demonstrate two things. First, feedback type training can be a valuable technique for improving clinical forecasts of a specific educational criterion. This technique may also prove useful for other training purposes such as the clinical interpretation of interests and personality inventories for certain kinds of cases, especially in building internal norms. Second, these results demonstrate the importance of considering

characteristics of the judges for whom training is considered. Some judges may profit from the training provided while others may not.

The results in this study are directed primarily toward the training of students. It is true that statistical predictions of institutional-type criteria are usually as accurate as the forecasts of the best judges. Therefore, in an actual prediction situation, serious consideration should be given to the desirability of attempting to train judges to achieve a level of accuracy already attained by the statistical method.

References

Cline, V. B. Ability to judge personality assessed with a stress interview and sound-film technique. Journal of Abnormal and Social Psychology, 1955, 50, 183-187.

Cronbach, L. J. Assessment of individual differences. In P. R. Farnsworth (Ed.), Annual Review of Psychology, Vol. VII. Stanford, Calif.: Annual Reviews, 1956. Pp. 173-196.

Crow, W. J. The effect of training upon accuracy and variability in interpersonal perception. Journal of Abnormal and Social Psychology, 1957, 55, 355-359.

Goldman, L. Using tests in counseling. New York: Appleton-Century-Crofts, 1961.

Gough, H. G. Clinical versus statistical prediction in psychology. In L. J. Postman (Ed.), Psychology in the making. New York: Knopf, 1962.

McArthur, C. Analyzing the clinical process. Journal of Counseling Psychology, 1954, 1, 203-208.

Meehl, P. E. Clinical vs. statistical prediction. Minneapolis: University of Minnesota Press, 1954.

Oskamp, S. The relationship of clinical experience and training methods to several criteria of clinical prediction. Psychological Monographs, 1962, 76 (28, Whole No. 547).

Sawyer, J. Measurement and prediction, clinical and statistical. Psychological Bulletin, 1966, 66, 178-200.

Soskin, W. F. Bias in postdiction from projective tests. Journal of Abnormal and Social Psychology, 1954, 49, 69-74.

Taft, R. The ability to judge people. Psychological Bulletin, 1955, 52, 1-23.

Taft, R. Multiple methods of personality assessment. Psychological Bulletin, 1959, 56, 333-352.

Tyler, L. E. The work of the counselor. New York: Appleton-Century-Crofts, 1961.

Watley, D. J. Counselor confidence in accuracy of predictions. Journal of Counseling Psychology, 1966, 13, 62-67. (a)

Watley, D. J. Counselor variability in making accurate predictions. Journal of Counseling Psychology, 1966, 13, 53-62. (b)

Watley, D. J. Do Counselors know when to use their heads instead of the formula? NMSC Research Reports, 1967, 3, No. 1.

PREVIOUS NMSC RESEARCH REPORTS

Volume 1, 1965

Number

1. The Inheritance of General and Specific Ability, by R. C. Nichols (also in Manosevitz, M., Lindzey, G., and Thiessen, D. (Eds.), Behavioral Genetics: Method and Research, Appleton-Century-Crofts, in press).

2. Personality Change and the College, by R. C. Nichols (also in American Educational Research Journal, 1967, 4, 173-190).

3. The Financial Status of Able Students, by R. C. Nichols (also in Science, 1965, 149, 1071-1074).

4. Progress of the Merit Scholars: An Eight-Year Followup, by R. C. Nichols and A. W. Astin (also in Personnel and Guidance Journal, 1966, 44, 673-686).

5. Prediction of College Performance of Superior Students, by R. J. Roberts.

6. Non-intellective Predictors of Achievement in College, by R. C. Nichols (also in Educational and Psychological Measurement, 1966, 26, 899-915).

7. Ninth Annual Review of Research by the NMSC Research Staff (superseded by the Tenth Annual Review).

8. Social Class and Career Choice of College Freshmen, by C. E. Werts (also in Sociology of Education, 1966, 39, 74-85).

Volume 2, 1966

1. Participants in the 1965 NMSQT, by R. C. Nichols.

2. Participants in the National Achievement Scholarship Program for Negroes, by R. J. Roberts and R. C. Nichols.

3. Career Choice Patterns: Ability and Social Class, by C. E. Werts.

4. Some Characteristics of Finalists in the 1966 National Achievement Scholarship Program, by W. S. Blumenfeld.

5. The Many Faces of Intelligence, by C. E. Werts (also in Journal of Educational Psychology, in press).

6. Sex Differences in College Attendance, by C. E. Werts.

7. Career Changes in College, by C. E. Werts (also in Sociology of Education, 1967, 40, 90-95).

Number

8. The Resemblance of Twins in Personality and Interests, by R. C. Nichols (also in Manosevitz, M., Lindzey, G., and Thiessen, D. (Eds.), Behavioral Genetics: Method and Research, Appleton-Century-Crofts, in press).

9. College Preferences of Eleventh Grade Students, by R. C. Nichols (also in College and University, in press).

10. The Origin and Development of Talent, by R. C. Nichols (also in Phi Delta Kappan, 1967, 48, 492-496).

11. Tenth Annual Review of Research, by the NMSC Research Staff (includes abstracts of all previous NMSC studies).

Volume 3, 1967

1. Do Counselors Know When to Use Their Heads Instead of the Formula?, by D. J. Watley (also in Journal of Counseling Psychology, in press).

2. Paternal Influence on Career Choice, by C. E. Werts (also in Journal of Counseling Psychology, in press).

NMSC research is supported by grants from the National Science Foundation, the Carnegie Corporation of New York, and the Ford Foundation